

zmPDSwR Chapter 1

coop711

2015년 9월 10일

Data

```
setwd("~/Dropbox/Works/Class/Data_Science/R.WD/zmPDSwR")
load("~/Dropbox/Works/Class/Data_Science/R.WD/zmPDSwR/GCDData.RData")
ls()
```

```
## [1] "creditdata"  "d"           "model"       "resultframe" "rtab"
## [6] "tab1"        "tab2"        "vars"
```

```
options(width=180)
str(creditdata)
```

```
## 'data.frame':    1000 obs. of  21 variables:
##  $ Status.of.existing.checking.account      : Factor w/ 4 lev
els "... < 0 DM", "... >= 200 DM / salary assignments for at least 1 year",...: 1
3 4 1 1 4 4 3 4 3 ...
##  $ Duration.in.month                      : int   6 48 12 42
24 36 24 36 12 30 ...
##  $ Credit.history                        : Factor w/ 5 lev
els "all credits at this bank paid back duly",...: 2 4 2 4 3 4 4 4 4 2 ...
##  $ Purpose                              : Factor w/ 10 le
vels "business", "car (new)",...: 8 8 5 6 2 5 6 3 8 2 ...
##  $ Credit.amount                        : int  1169 5951
2096 7882 4870 9055 2835 6948 3059 5234 ...
##  $ Savings.account.bonds                : Factor w/ 5 lev
els "... >= 1000 DM",...: 5 2 2 2 2 5 4 2 1 2 ...
##  $ Present.employment.since              : Factor w/ 5 lev
els "... >= 7 years",...: 1 3 4 4 3 3 1 3 4 5 ...
##  $ Installment.rate.in.percentage.of.disposable.income : int   4 2 2 2 3
2 3 2 2 4 ...
##  $ Personal.status.and.sex               : Factor w/ 4 lev
els "female : divorced/separated/married",...: 4 1 4 4 4 4 4 4 2 3 ...
##  $ Other.debtors.guarantors              : Factor w/ 3 lev
els "co-applicant",...: 3 3 3 2 3 3 3 3 3 3 ...
##  $ Present.residence.since               : int   4 2 3 4 4
4 4 2 4 2 ...
##  $ Property                             : Factor w/ 4 lev
els "if not A121 : building society savings agreement/life insurance",...: 3 3 3
1 4 4 1 2 3 2 ...
##  $ Age.in.years                         : int   67 22 49 4
5 53 35 53 35 61 28 ...
##  $ Other.installment.plans               : Factor w/ 3 lev
els "bank", "none",...: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Housing                             : Factor w/ 3 lev
els "for free", "own",...: 2 2 2 1 1 1 2 3 2 2 ...
##  $ Number.of.existing.credits.at.this.bank : int   2 1 1 1 2
1 1 1 1 2 ...
##  $ Job                                  : Factor w/ 4 lev
els "management/ self-employed/highly qualified employee/ officer",...: 2 2 4 2
2 4 2 1 4 1 ...
##  $ Number.of.people.being.liable.to.provide.maintenance.for: int   1 1 2 2 2
2 1 1 1 1 ...
##  $ Telephone                           : Factor w/ 2 lev
els "none", "yes, registered under the customers name": 2 1 1 1 1 2 1 2 1 1 ...
##  $ foreign.worker                       : Factor w/ 2 lev
els "no", "yes": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Good.Loan                           : Factor w/ 2 lev
els "BadLoan", "GoodLoan": 2 1 2 2 1 2 2 2 2 1 ...
```

- 작업 편의상 creditdata 를 등록

```
attach(creditdata)
head(Good.Loan, n=10)
```

```
## [1] GoodLoan BadLoan GoodLoan GoodLoan BadLoan GoodLoan GoodLoan GoodLoan
GoodLoan BadLoan
## Levels: BadLoan GoodLoan
```

```
table(Good.Loan)
```

```
## Good.Loan
## BadLoan GoodLoan
##      300      700
```

```
options(digits=2)
prop.table(table(Good.Loan))
```

```
## Good.Loan
## BadLoan GoodLoan
##      0.3      0.7
```

```
table(Credit.history, Good.Loan)
```

```
##
## Credit.history
##
## all credits at this bank paid back duly
1
## critical account/other credits existing (not at this bank)
3
## delay in paying off in the past
0
## existing credits paid back duly till now
1
## no credits taken/all credits paid back duly
5
```

| | Good.Loan | |
|--|-----------|---------|
| | BadLoan | GoodLoa |
| all credits at this bank paid back duly | 28 | 2 |
| critical account/other credits existing (not at this bank) | 50 | 24 |
| delay in paying off in the past | 28 | 6 |
| existing credits paid back duly till now | 169 | 36 |
| no credits taken/all credits paid back duly | 25 | 1 |

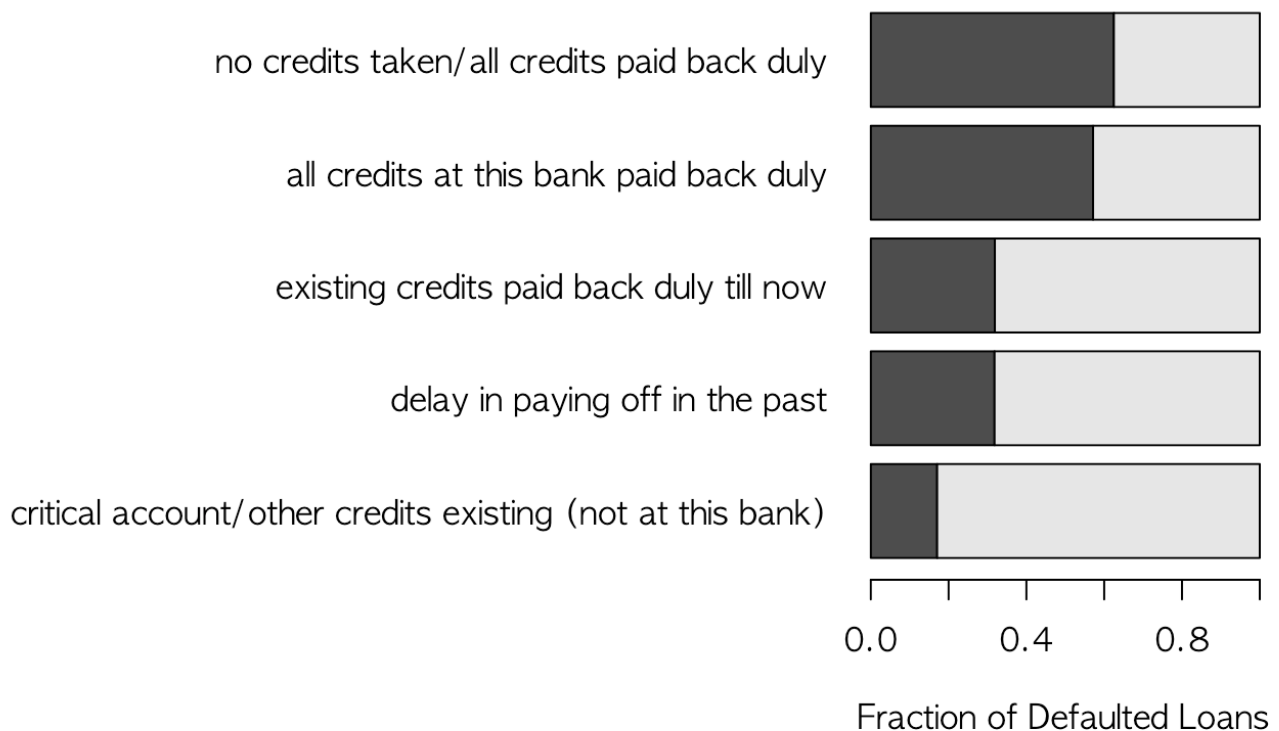
```
prop.table(table(Credit.history, Good.Loan), margin=1)
```

| | | Good.Loan | |
|----------------|--|-----------|---------|
| Credit.history | | BadLoan | GoodLoa |
| n | | | |
| ## | all credits at this bank paid back duly | 0.57 | 0.4 |
| 3 | | | |
| ## | critical account/other credits existing (not at this bank) | 0.17 | 0.8 |
| 3 | | | |
| ## | delay in paying off in the past | 0.32 | 0.6 |
| 8 | | | |
| ## | existing credits paid back duly till now | 0.32 | 0.6 |
| 8 | | | |
| ## | no credits taken/all credits paid back duly | 0.62 | 0.3 |
| 8 | | | |

```

par(pin=c(4, 4), mai=c(1.0, 4.5, 1.0, 0.5))
CG.percentage <- prop.table(table(Credit.history, Good.Loan), margin=1)
o.GL <- order(CG.percentage[,1])
barplot(t(CG.percentage[o.GL,]),las=1, horiz=TRUE)
title(xlab="Fraction of Defaulted Loans", ylab="")

```

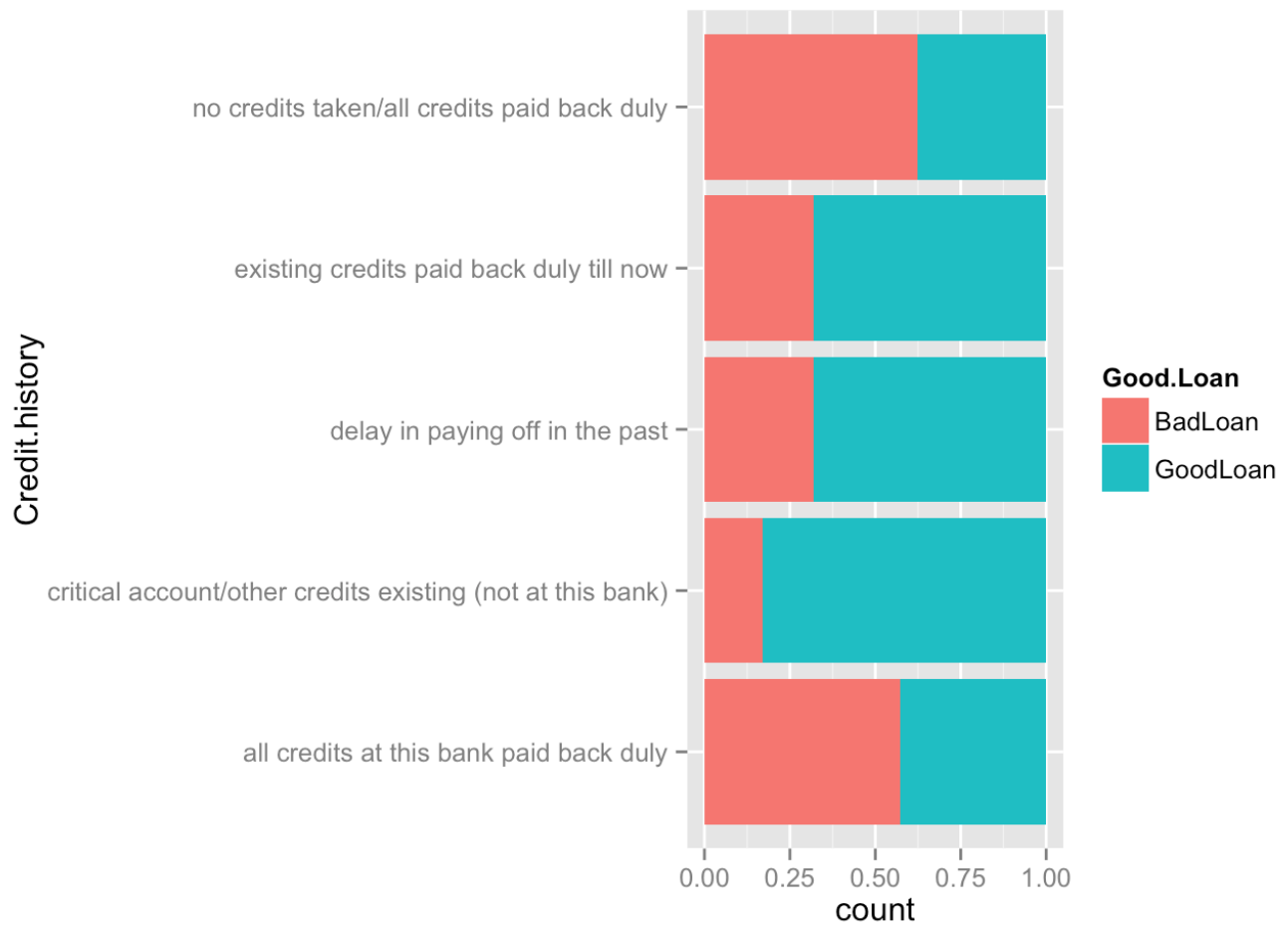


- ggplot 으로 그리면,

```

library(ggplot2)
ggplot(creditdata) + geom_bar(aes(x=Credit.history, fill=Good.Loan), position="fill") + coord_flip()

```



- Listing 1.1

```
library(rpart)
model <- rpart(Good.Loan ~ Duration.in.month +
               Installment.rate.in.percentage.of.disposable.income +
               Credit.amount +
               Other.installment.plans,
               data=d,
               control=rpart.control(maxdepth=4),
               method="class")
model
```

```
## n= 1000
##
## node), split, n, loss, yval, (yprob)
##      * denotes terminal node
##
## 1) root 1000 300 GoodLoan (0.30 0.70)
##    2) Duration.in.month>=34 170 82 GoodLoan (0.48 0.52)
##      4) Credit.amount< 2.2e+03 8 1 BadLoan (0.88 0.12) *
##      5) Credit.amount>=2.2e+03 162 75 GoodLoan (0.46 0.54)
##        10) Duration.in.month>=44 69 30 BadLoan (0.57 0.43)
##          20) Credit.amount< 7.4e+03 37 12 BadLoan (0.68 0.32) *
##          21) Credit.amount>=7.4e+03 32 14 GoodLoan (0.44 0.56) *
##        11) Duration.in.month< 44 93 36 GoodLoan (0.39 0.61) *
##    3) Duration.in.month< 34 830 220 GoodLoan (0.26 0.74)
##      6) Credit.amount>=1.1e+04 9 0 BadLoan (1.00 0.00) *
##      7) Credit.amount< 1.1e+04 821 210 GoodLoan (0.25 0.75) *
```

- Listing 1.2

```
resultframe <- data.frame(Good.Loan=creditdata$Good.Loan,
                          pred=predict(model, type="class"))
rtab <- table(resultframe)
rtab
```

```
##           pred
## Good.Loan  BadLoan GoodLoan
##   BadLoan      41      259
##   GoodLoan     13      687
```

```
sum(diag(rtab))/sum(rtab)
```

```
## [1] 0.73
```

```
sum(rtab[1,1])/sum(rtab[,1])
```

```
## [1] 0.76
```

```
sum(rtab[1,1])/sum(rtab[1,])
```

```
## [1] 0.14
```

```
sum(rtab[2,1])/sum(rtab[2,])
```

```
## [1] 0.019
```

- Listing 1.3

